

EXHIBIT F

9/19/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Todor Mihaylov, Ph.D.
Highly Confidential - Attorneys' Eyes Only - Under the Protective Order

Page 1

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

IN RE MATTER OF:)
RICHARD KADREY, et al.,)
Plaintiff,)
vs.) C.A. NO.:
META PLATFORMS, INC.,) 3:23-cv-03417-VC
Defendant.)
)

** HIGHLY CONFIDENTIAL - ATTORNEYS EYES ONLY **

UNDER THE PROTECTIVE ORDER

VIDEOTAPED DEPOSITION OF TODOR MIHAYLOV, Ph.D.

Palo Alto, California

Thursday, September 19, 2024

Stenographically Reported by:
HEATHER J. BAUTISTA, CSR, CRR, RPR, CLR
Realtime Systems Administrator
California CSR License #11600
Oregon CSR License #21-0005
Washington License #21009491
Nevada CCR License #980
Texas CSR License #10725

DIGITAL EVIDENCE GROUP
1730 M Street, NW, Suite 812
Washington, D.C. 20036
(202) 232-0646

9/19/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Todor Mihaylov, Ph.D.
Highly Confidential - Attorneys' Eyes Only - Under the Protective Order

Page 75

1 as one of his projects.

2 Q. Okay.

3 Is he still at Meta?

4 A. I don't know.

5 Q. Okay.

6 I was wondering if you could walk me
7 through this chat.

8 A. Yes.

9 So it seems here that Hugo is introducing
10 me to Damien --

11 Q. Um-hum.

12 A. -- who tries to generate MMLU, like,
13 questions from some knowledge text in different
14 categories.

15 Q. And so when you say "tries to demonstrate
16 some MMLU, like, knowledge questions," is the sample
17 below an example of that?

18 "What was Dr. Ann Wigmore's discovery
19 regarding her physical and mental state after
20 following a live food diet?"

21 A. Yes.

22 Q. So that's a question; right?

9/19/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Todor Mihaylov, Ph.D.
Highly Confidential - Attorneys' Eyes Only - Under the Protective Order

Page 76

1 A. Yes.

2 Q. And how is that question -- how has he
3 created that question? What is -- what is the role
4 of that sample in this chat?

5 A. So the role of this sample is, according to
6 my understanding, to show, like, how he generated
7 this. You can see that there is a prompt view --

8 Q. Um-hum.

9 A. -- which shows the prompt that he --

10 Q. Um-hum.

11 A. -- he used to ask, it seems LLaMA 2
12 model --

13 Q. Um-hum.

14 A. -- as it is mentioned, to generate that a
15 question and answer pairs.

16 (Stenographer clarification.)

17 THE WITNESS: Pairs.

18 Q. (By Ms. Poueymirou) So you're saying that
19 he's putting this prompt into LLaMA 2?

20 A. Yes, he probably did that.

21 Q. Okay.

22 And it's a prompt that says, "You are a

9/19/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Todor Mihaylov, Ph.D.
Highly Confidential - Attorneys' Eyes Only - Under the Protective Order

Page 77

1 helpful assistant that generates two multiple choice
2 questions." Is that the prompt you're referring to?

3 A. Yes, this is the prompt.

4 Q. Okay.

5 And then where you have the text that says
6 "generally I choose a light snack such as blending
7 apples," I assume that's "with avocado." What --
8 what is this text? Do you know?

9 A. No idea.

10 Q. Presumably a book? Who knows; right?

11 And so what is the relationship between
12 that prompt and text with the questions above with
13 these "she felt more exhausted and stressed to her
14 previous work schedule. She experienced no
15 difference" --

16 A. So I can't read the whole thing. It's too
17 much.

18 Q. Oh, no. So I -- above it. Above it.

19 A. Yes. So my understanding is --

20 Q. Yeah?

21 A. -- that in this prompt he asked the
22 model --

9/19/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Todor Mihaylov, Ph.D.
Highly Confidential - Attorneys' Eyes Only - Under the Protective Order

Page 78

1 Q. Um-hum.

2 A. -- to generate the questions and answers
3 and then probably the model generated it.

4 Q. So he puts a prompt in, plus some text from
5 a source we don't know, and then asks the model to
6 generate multiple-choice questions?

7 A. Yes, that's what it seems.

8 Q. And what is the purpose of that?

9 A. So the purpose is to use these questions
10 and answers as an evaluation for a model.

11 Q. It helps to determine how good the model
12 is?

13 A. Yes, in this topic.

14 Q. Okay.

15 And does that enable -- is the goal of that
16 to enable you to improve model performance?

17 A. So the goal is to evaluate the model for
18 this particular subject.

19 Q. And is the goal of evaluating models
20 ultimately to improve model performance?

21 A. Yes.

22 Q. Do you know what MMLIBU or M-M-L-I-B-U,

9/19/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Todor Mihaylov, Ph.D.
Highly Confidential - Attorneys' Eyes Only - Under the Protective Order

Page 79

1 would refer to?

2 A. I mean, yes, so it seems that it is a data
3 set generated with this particular --

4 (Stenographer clarification.)

5 THE WITNESS: -- method.

6 Q. (By Ms. Poueymirou) Do you have any idea
7 what "LIBU" refers to?

8 A. I don't. I think it might be LibGen.

9 Q. LibGen.

10 And what is LibGen?

11 A. LibGen is a library --

12 Q. Um-hum.

13 A. -- with data.

14 Q. Do you know what LibGen comprises?

15 A. Library Genesis, I think.

16 (Stenographer clarification.)

17 THE WITNESS: Library Genesis. This is
18 like how the name is formed.

19 Q. (By Ms. Poueymirou) The name, the full
20 name of LibGen is Library Genesis, yeah.

21 Do you know what is meant by a shadow
22 library?

9/19/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Todor Mihaylov, Ph.D.
Highly Confidential - Attorneys' Eyes Only - Under the Protective Order

Page 80

1 A. Which?

2 Q. Shadow library?

3 A. No.

4 Q. Do you know the kinds of texts that exist
5 in Library Genesis?

6 A. Yes.

7 Q. And what does Library Genesis comprise?

8 A. It comprises from books, scientific
9 articles, and magazines.

10 Q. Did Meta use LibGen as a -- as data for any
11 of its LLaMA models?

12 A. It -- we used the version of it that was
13 downloaded and processed.

14 Q. And do you know who downloaded that?

15 A. I think it was Nikolay Bashlykov.

16 Q. When you say "we used a version," what do
17 you mean by that?

18 A. I mean it was processed in some way.

19 Q. Do you mean it was filtered?

20 A. I don't know how it was processed.

21 Q. Okay.

22 Do you know if Meta still uses LibGen?

9/19/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Todor Mihaylov, Ph.D.
Highly Confidential - Attorneys' Eyes Only - Under the Protective Order

Page 131

1 What do you take that to mean?

2 MR. WEINSTEIN: Object to form.

3 THE WITNESS: So I can only interpret,
4 like, what is written here; that it is removing
5 codes that are containing copyright in the first or
6 the last 20 percent -- 25 percent of the book.

7 Q. (By Ms. Poueymirou) Did -- did you ever
8 process data and remove copyright from data when you
9 were preparing data?

10 A. I personally, no.

11 Q. Do you know folks that did and, if so, who?

12 A. Yeah, so it seems that Nikolay did that if
13 he is now the author of the document.

14 Q. Do you have any idea why Meta would want
15 copyright removed from books?

16 A. I don't know.

17 Q. Do you know how copyright would have been
18 removed from books?

19 MR. WEINSTEIN: Object to form.

20 THE WITNESS: I mean, I can guess what is
21 written here.

22 Q. (By Ms. Poueymirou) Yeah.